

Trabalho Final - Parte 1

João Vitor Yukio Bordin Yamashita

October 26, 2022

1 Contextualização e Aquisição

O objetivo desse projeto é criar um sistema de avaliação de risco cardíaco interativo e que, com base em dados, auxilie no diagnóstico de um paciente. A avaliação do risco cardíaco é a utilização de um grupo de exames para indicar a probabilidade de um evento cardiovascular, como um infarto do miocárdio ou um acidente vascular cerebral. A probabilidade é expressa em termos de grau de risco: pequeno, moderado e alto.¹

1.1 Entendendo o negocio

1.1.1 Quem vai interagir com o projeto?

A avaliação usa como um de seus parâmetros o perfil lipídico do paciente, que consiste em dados de colesterol e triglicerídeos, além de dados como pressão sanguínea e nível de açúcar no sangue. Com isso em mente as pessoas que irão interagir com o projeto seriam médicos e enfermeiros que coletarão e analisarão os dados.

1.1.2 Área do problema

Como mencionado anteriormente, o propósito da avaliação de risco cardíaco é identificar a possibilidade de um evento cardiovascular, uma avaliação cardiológica pode determinar um diagnóstico precoce desse tipo de doença. Quanto mais cedo um problema cardíaco for identificado, mais eficaz será o tratamento. Em muitos casos, a mudança de hábitos e de estilo de vida ajuda muito no tratamento, e é uma das primeiras orientações passadas pelo médico.

1.1.3 Estratégias existentes

A análise de risco cardíaco já é feita há um bom tempo, mas geralmente são usados métodos puramente manuais, como o "Assessing Cardiovascular Risk Guidelines"², que pode apresentar dificuldades em termos de comparação com outros pacientes e não provê nenhum tipo de interação com os dados do paciente. Além disso temos algoritmos como escore de risco de Framingham (ERF)³ e o escore de Reynolds que acrescenta a proteína Creatina (PCR). Esses algoritmos usam faixas de valores para determinar uma probabilidade de se ter a probabilidade de possuir um acidente cardiovascular. Esse projeto não tenta substituir os testes já utilizados, mas sim, prover uma ferramenta que possibilite uma exploração maior dos dados do paciente e encontrar casos que fogem do padrão.

1.2 Objetivos

1.2.1 Objetivos principais

O principal objetivo desse projeto é conseguir fornecer uma ferramenta que possibilite a análise e exploração dos dados com a finalidade de encontrar anomalias e características dos dados usados na avaliação de risco cardíaco de um paciente. Além disso, busca-se ter uma forma baseada em dados para

¹<https://labtestsonline.org.br/tests/avaliacao-de-risco-cardiaco>

²<https://www.nhlbi.nih.gov/sites/default/files/media/docs/risk-assessment.pdf>

³<https://linhasdecuidado.saude.gov.br/portal/obesidade-no-adulto/unidade-de-atencao-primaria/planejamento-terapeutico/escore-risco-global-framingham/>

avaliar o risco do paciente, que pode ser atualizada e não fixa, como é o caso do método mencionado anteriormente.

1.2.2 Questões a serem respondidas

A principal questão a ser respondida com o uso desse projeto é se um paciente, com base nos dados coletados, possui alguma tendência em termos de doença cardíaca quando comparado com outros dados de outros pacientes?

1.2.3 Como julgar o sucesso do projeto?

Se for possível classificar os dados de um paciente, ou, encontrar padrões nos dados que indiquem possíveis sinais de doenças cardíacas com base em modelos de IA ou com base em clusterização dos próprios dados, isso pode ser considerado um sucesso no auxílio da análise do paciente com base nos dados.

1.3 Entendimento da situação

1.3.1 Recursos disponíveis

Em termos de *hardware*, a ideia seria criar uma página web, que seja interativa para a análise dos dados, logo um computador com navegador web seria o necessário para o uso do projeto, para casos de uso de inteligência artificial e técnicas de *deep learning* podemos usar de GPUs disponíveis no Google Colab.

Para software, a ideia é usar o OpenRefine⁴ para limpeza dos dados, usar Python e a biblioteca Dash⁵ para criar o dashboard, além disso, com a finalidade de fornecer uma segunda opinião, a ideia é criar modelos de IA usando Scikit⁶ e TensorFlow⁷ para analisar os dados.

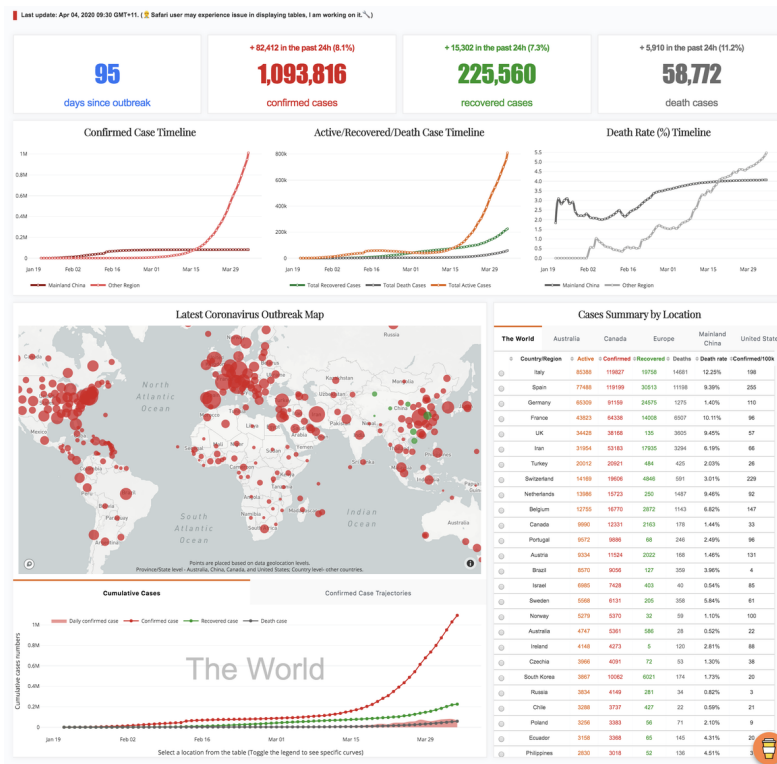


Figure 1: Exemplo de dashboard criado usando dash

⁴<https://openrefine.org/>

⁵<https://dash.plotly.com/>

⁶<https://scikit-learn.org/stable/>

⁷<https://www.tensorflow.org/>

Temos a seguir um exemplo de dashboard criado para análise de torres de celular no mundo (<https://dash.gallery/dash-world-cell-towers/>), além de outros exemplos disponíveis (<https://dash.gallery/Portal/>).

E por fim, na parte de recursos humanos, seriam os dados, que no caso, serão usados datasets disponíveis no Kaggle, como Cardiovascular Disease dataset⁸ e dados disponibilizados pela UCI⁹.

1.3.2 Quais são os dados?

Como mencionado anteriormente, a ideia seria usar datasets disponibilizados de forma online, o dataset do Kaggle possui 70000 amostras, caso seja observado que faltam dados, seria necessário agregar esses dados com outros datasets, como o da UCI mencionado anteriormente.

1.3.3 Restrições e riscos

Primeiramente, como se trata de um projeto envolvendo saúde, ele não deverá ser levado como verdade absoluta e sempre deverá ser usado como ferramenta de auxílio no diagnóstico. Além disso, caso seja feita a coleta de dados no local, é necessário disponibilizar uma forma de incluir (caso o paciente esteja de acordo) os dados ao banco de dados, ou, para utilizar os dados de forma anônima.

⁸<https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>

⁹<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>